

Published in final edited form as:

*Clin Lab Med.* 2018 September ; 38(3): 499–513. doi:10.1016/j.cll.2018.05.004.

## Harmonization of Liquid Chromatography–Tandem Mass Spectrometry Protein Assays

Alan L. Rockwood, PhD, DABCC<sup>a,\*</sup>, Mark S. Lowenthal, PhD<sup>b</sup>, Cory Bystrom, PhD<sup>c</sup>

<sup>a</sup>Rockwood Scientific Consulting, 11778 Stone Hollow Court, Salt Lake City, UT 84065, USA;

<sup>b</sup>Material Measurement Laboratory, Biomolecular Measurement Division, National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 6315, Gaithersburg, MD 20899, USA;

<sup>c</sup>Research and Development, Cleveland HeartLab, 6701 Carnegie Avenue, Suite 500, Cleveland, OH 44103, USA

### Keywords

LC-MS/MS; Protein; Harmonization; Standardization

## INTRODUCTION

Harmonization of diagnostic test results is fundamental to the effective use of laboratory testing in the diagnosis, treatment, and monitoring of disease. Working in an environment without any effort for diagnostic test harmonization might lead to diagnostic and therapeutic mistakes.

The International Consortium for Harmonization of Clinical Laboratory Results, convened in 2010, published a position statement<sup>1</sup> that defined 2 concepts; standardization (“uniformity of test results based on relation to a reference method”) and harmonization (“uniformity of test results when a reference method is not available”). Although this statement is recent, it recognizes and elevates an old problem in laboratory medicine, the need for laboratory measurements to be equivalent within agreed and meaningful limits.

Several laboratory tests that have population-wide impact on human health have undergone this process (eg, cholesterol, glucose, hemoglobin A1c); however, few if any mass spectrometry–based methods have reached a level of harmonization or standardization presented in the American Association for Clinical Chemistry (AACC) position statement. In this respect, mass spectrometry (MS) is not unique because relatively few tests in the clinical laboratory have undergone the rigorous harmonization process advocated in this document. The present article discusses some of the issues relevant to MS-based assay harmonization and standardization with a focus on proteins.

\*Corresponding author. alan@rockwood-scientific-consulting.com.

Disclosures: A.L. Rockwood and M.S. Lowenthal have no conflicts to disclose. C. Bystrom is an employee of Cleveland HeartLab.

## APPROACHES TO HARMONIZATION AND STANDARDIZATION

Harmonization and standardization is a formal effort among a wide range of stakeholders that start with the definition of a clinically relevant measurand. Subsequently, measurement methods are obtained (or developed) and evaluated for their ability to reproducibly determine the measurand in patient samples. For methods that will ultimately be standardized, reference methods and materials are developed in parallel so that traceability can ultimately be achieved. A roadmap for this process has been described.<sup>2</sup> Key components of this approach are illustrated in Figs. 1 and 2.

Standardization takes the concept of harmonization to a higher level: methods are not just harmonized with each other, but also with an agreed-on absolute standard of accuracy. Based on concepts outlined in ISO 17511:2003, a fully standardized method includes a traceability chain that is often presented in the form of a diagram mapping out a hierarchy of materials and procedures providing traceability of results back to a primary standard (Fig. 3).

The top of the traceability chain begins with definition of SI units for a selected measurand. On the left-hand side of the ladder we have derived materials used for measurement, while on the right-hand side we have a series of measurement methods. At each step of the ladder from top to bottom, materials and methods are fully evaluated so that trueness and imprecision can be accounted for. From a practical point of view, routine laboratory measurements at the bottom of the hierarchy are typically performed in high volumes and with the largest uncertainty, but under this scheme the measurement can be ensured not to exceed an established error budget.

The schemes summarized in Figs. 1–3 represent an aspirational goal for fully harmonized and/or standardized methods, although relatively few methods in clinical chemistry have completed this rigorous process to date, and to our knowledge no routine MS-based methods are among these fully harmonized or standardized methods, although in the small-molecule realm, significant progress toward that goal has been made by laboratories that have voluntarily participated in the Centers for disease Control and Prevention's (CDC's) Hormone Standardization (HoSt) Program<sup>3</sup> and Vitamin D Standardization-Certification Program (VDSCP).<sup>4</sup>

Although the roadmap provides a starting point for planning a formal harmonization project, it makes evident that this is a large-scale project best suited to commercially produced assays from multiple vendors, and such efforts may take many years or even decades to bear fruit, but more informal harmonization investigations carried out in laboratories before full-scale harmonization or standardization efforts also have value.

## DEFINING THE MEASURAND

Seldom, if ever, does a protein exist in a single form. Rather numerous proteoforms encompass protein variability that arise in the process of transcription, translation, and posttranslational modification. For harmonization efforts to be clinically relevant, it is essential that the appropriate proteoform(s) be targeted by the analytical method. Defining the appropriate measurand requires extensive research and can impact harmonization of

liquid chromatography–tandem MS (LC-MS/MS) methods in multiple ways, some favorable and some unfavorable.

MS-based detection is highly capable of differentiating heterogeneous proteoforms. These methods may be general, targeting a wide variety of proteoforms, or more narrowly targeted on a subset of proteoforms. Regardless of which proteoforms are selected, they should be relevant to the clinical needs.

Brain natriuretic peptide (BNP) is a good example. It is secreted as a 108-amino acid prohormone (proBNP) and is useful as a biomarker to rule out acute heart failure in an emergency setting. However, harmonization of current immunoassay platforms for BNP are problematic due to variability of cross-reactivity for different forms of BNP for assays from different vendors. The variable analytical specificity of immunoassay platforms is one limiting factor to establishing harmonized BNP results.<sup>5</sup> MS approaches may be more capable of addressing the complex metabolism of the natriuretic peptide family, and identify diagnostically relevant forms.

## REAGENTS FOR HARMONIZATION

A foundation for harmonization is agreement on an appropriate material against which test methods can be compared. From a bioanalytical perspective, the ideal material is a fully characterized pure protein, biochemically and biophysically identical to the native protein, which retains these same properties when added to a realistic matrix in required proportions. This ideal situation would allow for highly accurate and reproducible manufacture of reference material.

Achieving this goal is exceptionally challenging. Reference material manufactured using recombinant protein can have limitations, often lacking natural posttranslational modifications (phosphorylation, glycosylation), and recombinant proteins may lack the correct tertiary or quaternary structure.

As an alternative approach, purified native proteins and/or native matrix can arguably have better agreement in terms of biochemical and structural features, but biological variability arising from donor pool differences and matrix-dependent challenges in purification can result in unwanted lot-to-lot variability. Fully native matrix also presents challenges in how to achieve controlled differences in levels if required.

In practice, there are 2 broad types of materials used for harmonization of protein assays: recombined materials prepared with highly purified naturally occurring or recombinant protein added to measurand-free matrix, or fully native matrix in which the measurand has been highly characterized. The National Institute for Biological Standards and Control (NIBSC), European Joint Research Center (JRC), National Institute of Standards and Technology (NIST), National Metrology Institute of Japan, and others provide certified standards and reference materials for a wide range of clinically important protein and peptide analytes (eg, IGF-1, NIBSC code 02/254; Insulin, NIBSC code 83/500; ApoA-I, JRC BCR-393; C-reactive protein, NIST SRM 2924; cTnI, NIST SRM 2921).

In any case, 2 challenges that must always be addressed when harmonizing against a reference material are how to assign a value to the reference material and how to ensure commutability of the reference material. These issues must be addressed when setting up and executing a harmonization project.

## MASS SPECTROMETRY AS A PREFERRED PLATFORM FOR HARMONIZATION AND STANDARDIZATION

MS is a platform especially well-suited for assay harmonization, both with regard to reference methods within a standardization program and as routine methods used for routine analysis of patient samples. First, although multiple vendors supply the market with a range of instruments with varying performance, each system operates from the same first principles. The physics of generating, focusing, filtering, and transmitting ions for detection use similar principles across vendors, which affords much similarity between instruments. This contrasts with other common technologies, in which detection and instrument systems can vary dramatically for the same analyte (eg, turbidometry, spectrophotometry, fluorometry, electrochemiluminescence, and chemiluminescence).

Second, MS relies on direct detection of specific molecules or molecular fragments based on a well-defined and easily understood physical property, that is, mass. In this case, the identity of the species used for detection and quantification can be determined using well-established principles and procedures, such as interpretation of tandem mass spectra and accurate mass measurements, either by de novo methods or by comparison against known standards. By way of contrast, detection in immunometric systems relies on the recognition of an epitope by an antibody, which is at heart a problem in natural product chemistry and is therefore subject to much uncertainty. Despite the ability to produce antibodies with high avidity and selectivity, the tremendous complexity of a matrix like human serum, as well as the relative irreproducibility of antibody production and characterization, makes guarantee of consistent fidelity challenging. Finally, the ability to incorporate heavy isotopes to prepare labeled internal standards substantially improves assay accuracy and precision in MS as well as providing insight into assay performance and quality control, which is difficult to replicate using other measurement techniques.

The combination of these advantages is often applied to the development of reference methods using isotope dilution (ID) techniques with LC-MS/MS. A common approach for harmonizing assays is to quantify protein measurands through a “bottom-up” workflow by detection of surrogate peptides that result from a proteolytic digestion of the clinically relevant protein. ID requires the use of heavy-isotope–labeled internal standards spiked into biological samples and into calibration standards, facilitating high-precision, low-bias quantification. Multiple-reaction monitoring (MRM) MS using a triple-quadrupole instrument is a highly selective technique that allows simultaneous, multiplexed detection of natural and heavy (ie, labeled) forms of surrogate peptides detected by precursor-to-fragment ion transitions. Whereas synthetic labeled peptides are appropriate for use in some applications, other harmonization efforts will require full-length recombinantly produced protein internal standards, a manufacturing challenge in itself.

The general strategy for protein quantification meant to underpin harmonization efforts has been established for LC-MS techniques. Yet to date, only a few reference measurement methods (RMs) for proteins have been developed using ID LC-MS/MS techniques. These internationally accepted RMs include clinical markers, such as amyloid beta 1 to 42, HbA1c, and C-peptide. As MS has become more widely available, the possibility of deploying routine methods with analytical performance close to that expected from reference methods has become possible. Despite the analytical utility of MS, development of methods with exemplary performance requires extensive knowledge, extensive development, and exhaustive validation.

## COMMUTABILITY AND MATRIX EFFECTS

A key requirement in any standardization program is that the reference materials must be analytically equivalent to the target compounds in patient samples, or in other words, commutability is a key issue, and matrix effects must be minimal or absent. We often think of poor commutability as primarily affecting immunoassays, but one must not overlook the fact that it can also affect MS-based assays. To illustrate this with 2 examples, electrospray ionization, which is widely used in the clinical laboratory, is subject to matrix-dependent suppression of ionization efficiency. To a large extent, the use of internal standards can overcome this issue, but if ion suppression is too severe, the signal levels may become too low to be usable. As a second example, it has become generally accepted that one should use ratios of ion abundances from different MS/MS transitions to detect the presence of interferences. Briefly, when this ratio for a patient sample differs from that of a pure standard, an interfering compound is deemed likely. However, experiments have shown cases in which the branching ratios for different MS/MS pathways can be matrix dependent.<sup>6,7</sup> In any standardization program, even those targeted to MS-based methods, the materials and methods must be rigorously evaluated for commutability and matrix effects.<sup>8–10</sup>

## EXAMPLES

This section discusses examples of both preliminary informal harmonization efforts and more formal harmonization efforts. Thyroglobulin (Tg) and insulin-like growth factor-1 (IGF-1) both represent important clinical markers that attracted interest in de novo test development by LC-MS/MS. Tg is a marker for thyroid cancer recurrence. After thyroidectomy, circulating levels of Tg are expected to decline to undetectable levels unless there is recurrence of disease. A recognized limitation of the immunometric measurement of Tg is the high proportion of patients who have circulating autoantibodies directed against Tg.

Measurement of IGF-1 is used in the diagnosis of growth disorders. This protein measurand circulates as a complex with IGF binding proteins. Much like autoantibodies, the presence of binding proteins makes accurate and precise quantification of IGF-1 by immunoassay difficult.

First let us consider Tg as a model for an informal preliminary approach to harmonization. It should be noted at the outset that the Tg work referenced here<sup>11,12</sup> was essentially a method

comparison study presented as a first step toward harmonization, and not as a full harmonization project, either formal or informal. For example, that study did not include the process of selecting an agreed-on reference material or reference method. Furthermore, we recognize that reasonable people may disagree on whether informal/preliminary harmonization efforts should be pursued at all.

Within the past decade, methods for Tg analysis by LC-MS/MS were under independent development by several laboratories. Within a relatively short time, several laboratories introduced their versions of Tg analysis by LC-MS/MS. Although similar in certain broad features (each using a surrogate peptide from a tryptic digest, for example), the method details varied substantially. To provide the best patient care possible, several of these laboratories, which were otherwise strongly competitive with each other, agreed to participate in a multi-laboratory method comparison study and to publish the results.

Fig. 4 summarizes the results from these studies. Necessarily, each laboratory had to initially pick some method or reference material to which to anchor their results.

The graph on the left presents the results for 4 LC-MS/MS methods compared with the average of the 4 methods. The graph on the right compares 4 well-established immunoassays. It is immediately obvious that the LC-MS/MS methods compare with each other at least as well as the immunoassays compare with each other, and probably better. Of particular relevance, compared with the immunoassays-based methods, the LC-MS/MS methods showed better agreement with each other when autoantibodies were present, which was the primary motivation for developing the LC-MS/MS methods to begin with. The agreement among the 4 LC-MS/MS methods is remarkable considering the detailed differences between the LC-MS/MS procedures described in the article.

In a similar vein, measurement of IGF-1 was evaluated among 4 international laboratories that independently developed LC-MS/MS methods and subsequently examined the consequences of various calibration strategies, which included the use of an NIBSC reference material. Similar to the findings for Tg, careful de novo assay development yielded intra-laboratory agreement that was better than observed using a commercial platform in 2 different laboratories.

One can take important lessons from these 2 examples. First, even though there were no established LC-MS/MS-based methods for comparison when the assays were under development, the potential for good harmonization was baked into them via choice of calibrators, as described in the article, as well as certain advantages inherent in MS-based methods, as discussed in the present article.

A second lesson from Tg is that given that these 4 assays were all being offered for patient care, it was important not to wait for a larger-scale harmonization program to be implemented before beginning some kind of harmonization effort. Part of the background to this is that the MS-based methods were developed to improve patient care (ie, deal with the problem of interference by autoantibodies in immunoassays), and in light of this it seemed inappropriate to wait years for a large-scale harmonization project before putting these assays into service.

Third, the study investigators explicitly recognized that this was a preliminary effort and that higher-level harmonization studies should be done.

Fourth, even though these laboratories operate in a competitive environment, all recognized the importance of working together in this area to provide the best patient care possible.

The approach used for the preliminary Tg harmonization can serve a model for other LC-MS/MS methods of protein analysis. Elements of this approach, some of which were foreshadowed in the last paragraph, include the following: (1) Although all laboratories may have a hand in structuring the study, it is important to have 1 person at 1 institution assigned to lead the process and to keep the process on track. (2) The classifications of samples used in the study should address potential issues or potential weaknesses relevant to a method. For example, the Tg study included a group of samples for which autoantibodies against Tg were present. This issue had already been identified as a possible weakness for immunoassays, and given that the MS-based method purported to overcome interferences arising from autoantibodies, it was important to address this issue in a harmonization study over multiple laboratories. (3) Performance parameters to be compared must be agreed on. Obviously, this would normally include comparison of quantitative results, but depending on the situation may include other parameters, such as lower limits of quantification. In the case of the Tg study, the lower limits of quantification were not compared, although the laboratories were transparent with regard to that specification. (4) The study members must agree to the number of samples to be used, what will be the sources of the samples, and the procedures used to share the samples. (5) In many cases, laboratory competition is an underlying issue that needs to be dealt with. Ideally, all laboratories offering a particular protein LC-MS/MS assay will agree to be involved in the study. In the case of Tg, most but not all of the laboratories offering the test at that time participated in the study. Furthermore, in the interest of transparency, the participating laboratories agreed to participate in the study nonanonymously and to have the results published in the open literature. Relevant to the competitiveness, intellectual property issues may sometimes make cooperation difficult, but for the good of the patient, it is important to try to work these issues out. (6) The participants should be aware that an early-phase method comparison is just a step toward the ultimate goal of a high-level harmonization program, or in the best case, a full standardization program, while at the same time recognizing that these higher-level harmonization goals may be years away.

A recent publication<sup>13</sup> illustrates a more formalized example of approaches to harmonization. This article is highly recommended for its discussion of formal harmonization processes applied to the real-life example of C-peptide. Among other things, that article provides a historical narrative of efforts to standardize C-peptide, starting in 2002 with the formation of a C-peptide Standardization Committee by the National Institute of Diabetes and Digestive and Kidney Diseases in the United States. The Diabetes Diagnostic Laboratory at the University of Missouri coordinated the effort that eventually became an international effort involving multiple agencies. The study was not focused specifically on standardization of MS-based end-user methods for C-peptide, but would apply to all methods for C-peptide. However, that work did eventually lead to an MS-based reference method for C-peptide. The development of a reference method has been an essential part of

the harmonization program, but in-and-of-itself does not complete the process of assay standardization, and the article lays out some steps to be taken in the future specific to the reference method itself.

This article draws several lessons, starting with the fact that someone has to pay for the work. Also noted is that communication among the many groups involved in the effort can be difficult. Other lessons are that the goals should be laid out at the beginning of the process, that the time frame for successful formal harmonization is long (more than a decade and a half so far in this example, and the standardization process is yet to be completed), and that the development of reference materials and procedures can be long, difficult, and expensive. The most fundamental lesson was left unstated in the article, which is that the process started with the fact that C-peptide assays were already being used in clinical practice. Thus, it was not a case of waiting for standardization to occur before an assay can be used for patient care. Indeed, given the level of effort required for full harmonization, it is not likely that any new assays would ever be introduced into practice if they must first undergo a full program of formal harmonization. The lessons learned from the C-peptide story are also applicable to less formalized harmonization efforts, differing mainly in degree and scope when applied to formal versus less-formal harmonization efforts.

## ROLE OF PROFICIENCY TESTING

Although this article has emphasized the need for more extensive harmonization efforts, one should not overlook the role of proficiency testing. It is best to think of proficiency testing programs not as full harmonization programs in and of themselves, but rather as programs to maintain harmony between methods within a harmonization effort. In many countries, proficiency testing programs have regulatory authority. If a laboratory consistently fails proficiency testing for a given analyte it may result in the laboratory being unable to offer that test.

Furthermore, as discussed later in this article, the CDC hormone standardization program includes a proficiency testing program for steroid hormones.<sup>14</sup> This could provide a model for the role of proficiency testing in harmonization and standardization efforts for protein assays by LC-MS/MS.

There are at least 3 aspects of proficiency testing that interact with each other to affect harmonization. One addresses the issue of whether methods themselves agree with each other (harmonization between methods). Another is whether different laboratories that run the same method produce results that agree with each other (harmonization within a peer group). The third is whether the methods agree with a well-defined and agreed-on standard of absolute accuracy (standardization). To understand the role of proficiency testing in practical harmonization efforts, one must keep these 3 aspects in mind.

One significant question that needs to be considered, even at the present stage of historical development, is whether it is sufficient to seek harmony between the methods in a proficiency testing program, or should the program aspire to the higher goal of absolute accuracy. An anecdote from the realm of LC-MS/MS testing of small molecules can help



frame the issue. Some years back, one of the laboratories performing LC-MS/MS testing of metanephrines was consistently failing proficiency testing within its peer group. That proficiency testing program was based on harmony between laboratories, not absolute accuracy. An investigation showed that the laboratory that was consistently failing proficiency testing was also the laboratory that was consistently producing the more accurate results.<sup>15</sup> Therefore, when constructing or participating in a proficiency testing program, it is important to consider the goals of the program. Is it to be harmony-based or accuracy-based, and how are the results to be evaluated to ensure the best patient care?

The commutability issue is also applicable in proficiency testing programs, particularly in cases in which samples may be unstable. In those cases, the proficiency samples may be designed to improve stability, which can sometimes have the side effect of making the samples less like natural patient samples and therefore at greater risk of being noncommutable.

## METHOD CONSIDERATIONS

In general, the methods used for analysis fundamentally affect the prospects for successful harmonization, and assays using LC-MS/MS for protein analysis are no different. As mentioned earlier, some MS methods for protein analysis have demonstrated good agreement even in the absence of rigorous use of reference materials. This is related to the fact that direct analysis of protein fragments broadens the options for first principle approaches to quantitative analysis.

In contrast to immunologic methods, in which access to selected epitopes is governed by immunoreactivity of the organism selected for antibody production, a diverse range of proteotypic peptides can be accessed based on the requirements of the method. In many cases, preanalytical conditions can be finely tuned to derive the measurement from peptides that demonstrate optimal properties for analysis, as well as containing specific amino acid polymorphisms, posttranslational modifications, or cleavages, as needed. This process is not without constraints. It is recognized that some proteins/peptides present biophysical and biochemical challenges that reduce their value for analytical characterization, and there remain proteins that are refractory to elements of the required preanalytical sample preparation.

A few things that should be considered when designing and validating a method include whether to use a top-down method (MS method of a whole protein, possibly including MS/MS), a bottom-up method (digesting the protein and detecting specific peptides as surrogates for the target protein), which peptides to target in a digestion-based method, how many peptides to target in the method, what calibrators to use, choice of internal standards, and the quality of the LC separation necessary for a successful LC-MS/MS-based method. All of these things ultimately affect the prospects for successful method harmonization. An example of a decision matrix for calibrator selection is given in Fig. 5.

## LITERATURE REVIEW/BIBLIOGRAPHY

Laboratory assays commonly provide results for 700 or more types of quantities, with varying degrees of metrological traceability.<sup>16</sup> Primary reference measurement procedures (RMPs) and primary reference materials are available for 25 to 30 (conservatively) types of quantities linking these assays traceable to the SI. Reference materials without associated RMPs can be found for more than 300 types of quantities. RMPs without primary reference materials are less common (w30 quantities). Regrettably, most current laboratory assays (>300) are run exclusively using “in-house” calibrators and measurement procedures.

Implementing a successful assay harmonization effort should be considered compulsory to ensure the comparability of patient results.<sup>17</sup> However, there are currently relatively few examples of successful harmonization efforts. That number is growing, largely reflecting the implementation of electronic health records by physicians and hospitals,<sup>18</sup> and thanks to government directives such as the European Union’s In Vitro Diagnostic Directive<sup>19</sup> or regulations such as the Food and Drug Administration’s Clinical Laboratory Improvement Amendments guidance.<sup>20</sup> Independent bodies, such as AACC/International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and Joint Committee for Traceability in Laboratory Medicine, World Health Organization (WHO), Clinical and Laboratory Standards Institute, as well as national entities like NIST and National Institutes of Health (NIH)/CDC, are leading the international harmonization effort. Included among the harmonization success stories to date are efforts such as the CDC’s HoSt program, which is focused on serum hormones like thyroid-stimulating hormone,<sup>21</sup> estradiol and, testosterone,<sup>22</sup> and the NIH/Office of Dietary Supplements Vitamin D Standardization Program.<sup>4,23</sup> Other organizations have also led efforts to create international standards for measurands like serum immunoglobulin E<sup>24</sup> (College of American Pathologists (CAP), WHO), free thyroxine,<sup>25</sup> triiodothyronine<sup>26</sup> (IFCC), and for lipid function tests such as cholesterol<sup>27</sup> (NIST), and kidney function measurement of serum creatinine<sup>28</sup> (NIST). In some cases, atypically, harmonization of clinical assays results from only a single manufacturer producing a test specific for a given measurand (N terminal-proBNP, cardiac troponin T)<sup>5,29</sup> (Roche Diagnostics, Risch-Rotkreuz, Switzerland).

The International Consortium for Harmonization of Clinical Laboratory Results provides an updated resource summarizing the measurand harmonization activities ongoing throughout the world (<http://www.harmonization.net/measurands>). Practical approaches to harmonizing routine laboratory methods have been described.<sup>30,31</sup> Useful resources are sufficiently available for understanding past and ongoing assay harmonization efforts, for establishing traceability of results, and for understanding commutability of standards.<sup>14</sup> Harmonization of laboratory testing requires cooperation from international stakeholder communities, including clinicians, medical device manufacturers, and standards, metrology, governmental, and professional organizations, and others.<sup>32</sup>

The CDC hormone standardization program provides a useful model that could be emulated for protein measurements by LC-MS/MS.<sup>33</sup> This program focuses on 4 main components: (1) developing and implementing reference methods, calibrated using “pure compound” hormones, (2) establishing an assay and laboratory calibration program to (among other

functions) ensure the calibration does not change over time, (3) working with proficiency testing companies to develop laboratory surveys to assess and improve the measurement of targeted hormones, and (4) collaborating with professional organizations and institutions to develop training and education materials.

## SUMMARY

Assay harmonization is a collaborative effort typically requiring long-term commitment, technical expertise, and financial resources. Achieving this significant goal is a worthwhile investment. As a practical matter, this article advocates a multistep approach to harmonization of LC-MS/MS-based protein assays, starting with small-scale preliminary harmonization efforts, culminating with full-scale harmonization efforts that follow the harmonization roadmap. The ultimate goal of fully standardized assays may remain elusive for quite some time. This approach will likely be done on an analyte-by-analyte basis in most cases, although there may sometimes be potential for multiplexing multiple.

LC-MS/MS is beginning to be applied to the analysis of proteins for diagnostic purposes. The high specificity of LC-MS/MS brings with it advantages and opportunities for improved analytical performance for improved diagnostic performance and better patient care; however, in some cases, this high degree of specificity can also be a risk if it targets clinically irrelevant forms. Proper assay development remains the key to avoiding this type of failure.

Harmonization between methods has received greater emphasis in recent years, and formalized approaches for method harmonization have been developed. However, relatively few clinical analytes have been harmonized or standardized to the high level required under these schemes. In this respect, MS-based methods are still at an early stage, possibly even lagging behind more established methods in the clinical laboratory. To our knowledge, no protein assays based on LC-MS/MS have been rigorously harmonized or standardized to date.

Nevertheless, as clinical laboratory scientists, we must diligently work toward harmonization of our LC-MS/MS methods, even if it is not yet practical to fully implement the most rigorous harmonization protocols that are being developed. We have presented the example of a preliminary Tg harmonization study as a prototype for bootstrapping the harmonization process; that is, early efforts to bring methods into harmonization before the implementation of the highest level of harmonization, and we have extracted lessons taken from that study and used them to outline a general approach for early-stage harmonization efforts.

Also presented are technical aspects of MS in relation to how they may affect the harmonization process and a short review of some of the relevant literature.

## ACKNOWLEDGMENTS

Helpful discussions with Geoffrey Rule are gratefully acknowledged.

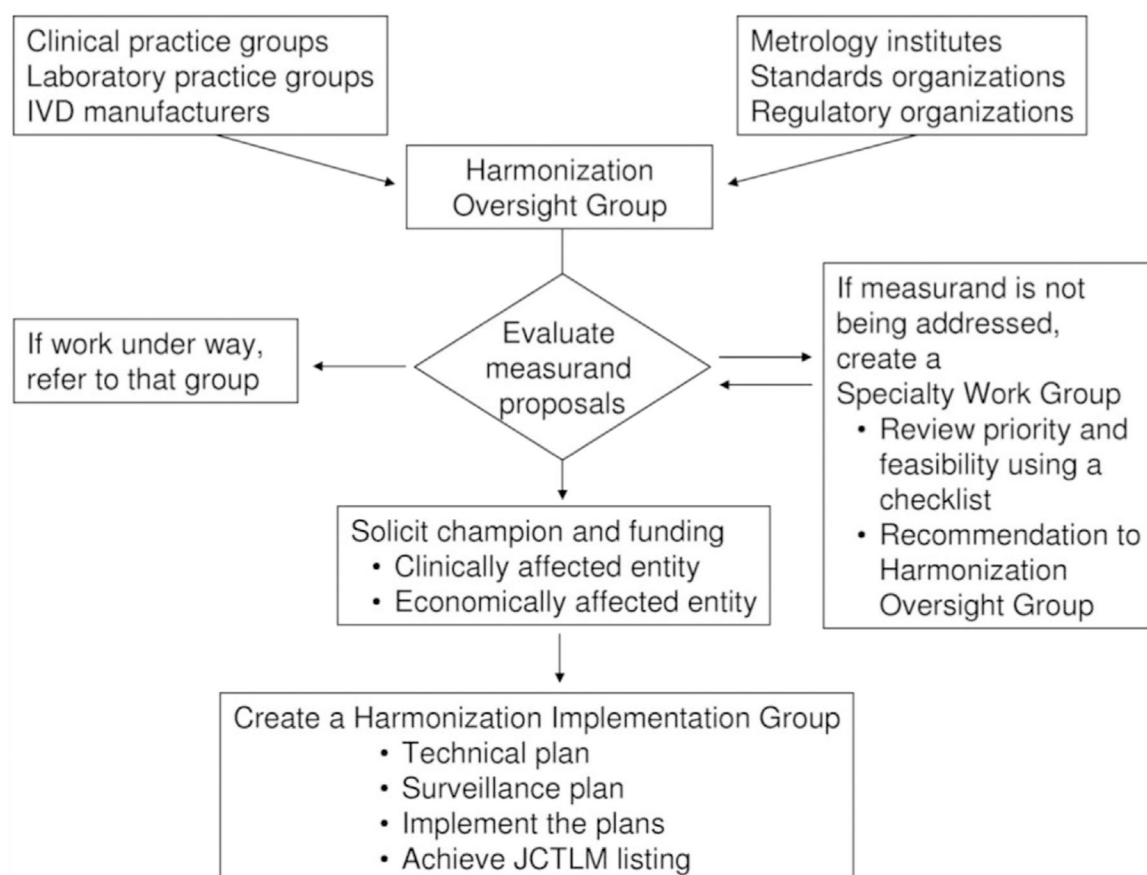
## REFERENCES

1. AACC POSITION STATEMENT - Harmonization of clinical laboratory test results [Web site], 2013 Available at: [https://www.harmonization.net/media/1087/aacc\\_harmonization\\_position\\_statement\\_2013.pdf](https://www.harmonization.net/media/1087/aacc_harmonization_position_statement_2013.pdf). Accessed June 7, 2018.
2. Greg Miller W, Myers GL, Lou Gantzer M, et al. Roadmap for harmonization of clinical laboratory measurement procedures. *Clin Chem* 2011;57(8):1108–17. [PubMed: 21677092]
3. Vesper HW, Botelho JC, Wang Y. Challenges and improvements in testosterone and estradiol testing. *Asian J Androl* 2014;16(2):178–84. [PubMed: 24407184]
4. Phinney KW, Semplos CT, Tai SS, et al. Baseline assessment of 25-Hydroxyvitamin D reference material and proficiency testing/external quality assurance material commutability: a vitamin D standardization program study. *J AOAC Int* 2017; 100(5):1288–93. [PubMed: 28797319]
5. Saenger AK, Rodriguez-Fraga O, Ler R, et al. Specificity of B-type natriuretic peptide assays: cross-reactivity with different BNP, NT-proBNP, and proBNP peptides. *Clin Chem* 2017;63(1):351–8. [PubMed: 28062628]
6. Kaufmann A, Butcher P, Maden K, et al. Are liquid chromatography/electrospray tandem quadrupole fragmentation ratios unequivocal confirmation criteria? *Rapid Commun mass Spectrom* 2009;23(7):985–98. [PubMed: 19241450]
7. Wang J, Aubry A, Bolgar MS, et al. Effect of mobile phase pH, aqueous-organic ratio, and buffer concentration on electrospray ionization tandem mass spectrometric fragmentation patterns: implications in liquid chromatography/tandem mass spectrometric bioanalysis. *Rapid Commun mass Spectrom* 2010;24(22): 3221–9. [PubMed: 20972995]
8. Nilsson G, Budd JR, Greenberg N, et al. IFCC working group recommendations for assessing commutability part 2: using the difference in bias between a reference material and clinical samples. *Clin Chem* 2018;64(3):455–64. [PubMed: 29348165]
9. Miller WG, Schimmel H, Rej R, et al. IFCC working group recommendations for assessing commutability part 1: general experimental design. *Clin Chem* 2018; 64(3):447–54. [PubMed: 29348163]
10. Budd JR, Weykamp C, Rej R, et al. IFCC working group recommendations for assessing commutability part 3: using the calibration effectiveness of a reference material. *Clin Chem* 2018;64(3):465–74. [PubMed: 29348164]
11. Netzel BC, Grebe SK, Carranza Leon BG, et al. Thyroglobulin (Tg) testing revisited: Tg assays, Tgab assays, and correlation of results with clinical outcomes. *J Clin Endocrinol Metab* 2015;100(8):E1074–83. [PubMed: 26079778]
12. Netzel BC, Grant RP, Hoofnagle AN, et al. First steps toward harmonization of LC-MS/MS thyroglobulin assays. *Clin Chem* 2016;62(1):297–9. [PubMed: 26430076]
13. Little RR, Wielgosz RI, Josephs R, et al. Implementing a reference measurement system for c-peptide: successes and lessons learned. *Clin Chem* 2017;63(9):1447–56. [PubMed: 28646033]
14. Prevention CfDCa. HoSt/VDSCP: standardization of measurement procedures. Laboratory Quality Assurance and Standardization Programs. 2017.
15. Singh RJ, Grebe SK, Yue B, et al. Precisely wrong? Urinary fractionated metanephrines and peer-based laboratory proficiency testing. *Clin Chem* 2005; 51(2):472–3 [discussion 473–4]. [PubMed: 15681566]
16. International Organization for Standardization (ISO). In vitro diagnostic medical devices – Measurement of quantities in biological samples – metrological traceability of values assigned to calibrators and control materials. ISO 17511:2003.
17. Panteghini M Traceability, reference systems and result comparability. *Clin Biochem Rev* 2007;28(3):97–104. [PubMed: 17909614]
18. Office-based Physician Electronic Health Record Adoption. Health IT Quick-Stat #50 2016; Health IT Quick-Stat #50. Available at: [dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php](https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php). Accessed June 7, 2018.
19. Directive 98/79/EC of the European parliament and of the council of 27 October 1998 on in vitro diagnostic medical devices. *Official Journal of the European Communities* 1998;41(L 331).

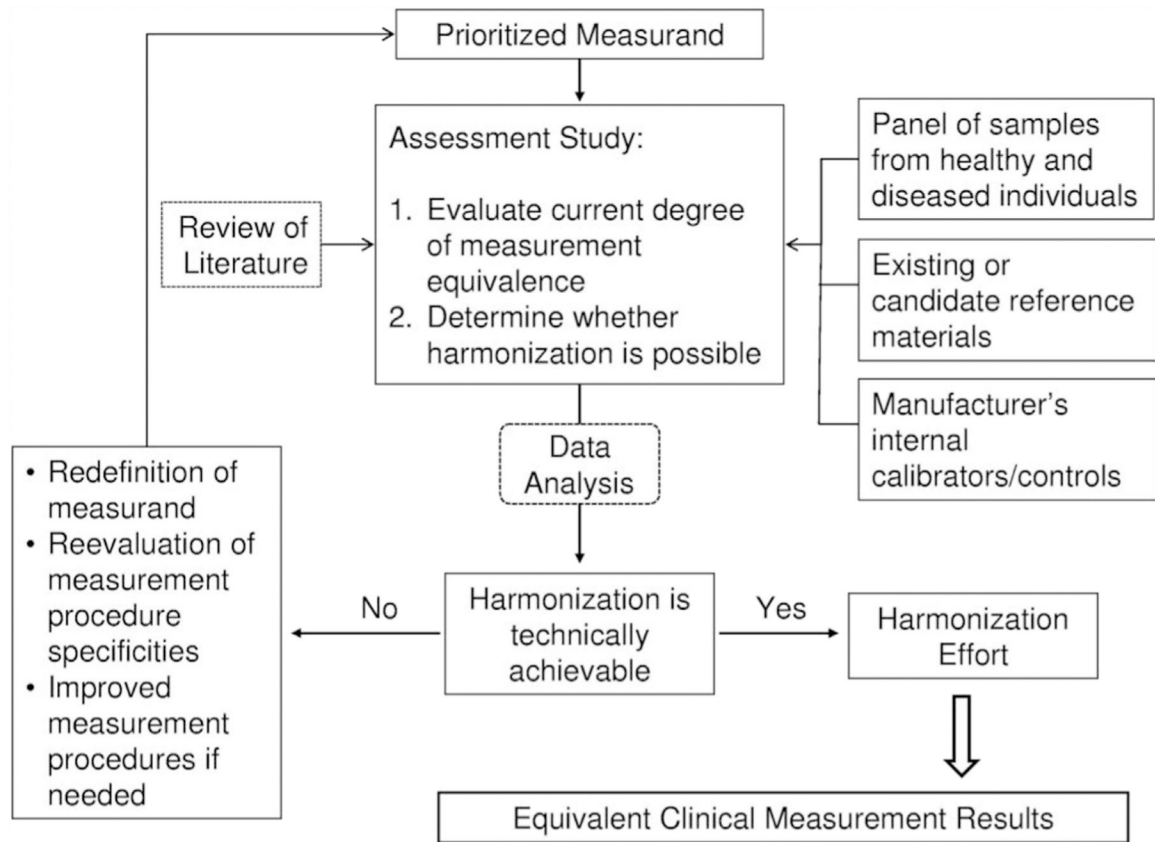
20. US Food and Drug Administration (FDA), 2018 Clinical Laboratory Improvement Amendments (CLIA), Medical Devices, IVD Regulatory Assistance. Available at: <https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm124105.htm>. Accessed June 7, 2018.
21. Thienpont LM, Van Uytanghe K, De Grande LAC, et al. Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. *Clin Chem* 2017;63(7): 1248–60. [PubMed: 28522444]
22. Travison TG, Vesper HW, Orwoll E, et al. Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the United States and Europe. *J Clin Endocrinol Metab* 2017;102(4):1161–73. [PubMed: 28324103]
23. Wise SA, Phinney KW, Tai SS, et al. Baseline assessment of 25-Hydroxyvitamin D assay performance: a Vitamin D standardization program (VDSP) interlaboratory comparison study. *J AOAC Int* 2017;100(5):1244–52. [PubMed: 28822355]
24. Thorpe SJ, Heath A, Fox B, et al. The 3rd International Standard for serum IgE: international collaborative study to evaluate a candidate preparation. *Clin Chem Lab Med* 2014;52(9):1283–9. [PubMed: 24807170]
25. Faix JD, Miller WG. Progress in standardizing and harmonizing thyroid function tests. *Am J Clin Nutr* 2016;104(Suppl 3):913s–7s. [PubMed: 27534642]
26. Thienpont LM, Van Uytanghe K, Beastall G, et al. Report of the IFCC working group for standardization of thyroid function tests; part 2: free thyroxine and free triiodothyronine. *Clin Chem* 2010;56(6):912–20. [PubMed: 20395623]
27. Ellerbe P, Meiselman S, Sniegowski LT, et al. Determination of serum cholesterol by a modification of the isotope dilution mass spectrometric definitive method. *Anal Chem* 1989;61(15):1718–23. [PubMed: 2672877]
28. Camara JE, Lippa KA, Duewer DL, et al. An international assessment of the metrological equivalence of higher-order measurement services for creatinine in serum. *Anal Bioanal Chem* 2012;403(2):527–35. [PubMed: 22392375]
29. Januzzi JL, van Kimmenade R, Lainchbury J, et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study. *Eur Heart J* 2006;27(3):330–7. [PubMed: 16293638]
30. Greaves RF. A guide to harmonisation and standardisation of measurands determined by liquid chromatography–tandem mass spectrometry in routine clinical biochemistry. *Clin Biochem Rev* 2012;33(4):123–32. [PubMed: 23267244]
31. Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardization and harmonization of clinical laboratory tests. *Am J Clin Nutr* 2016; 104(Suppl 3):907s–12s. [PubMed: 27534625]
32. Tate JR, Johnson R, Barth J, et al. Harmonization of laboratory testing—current achievements and future strategies. *Clin Chim Acta* 2014;432:4–7. [PubMed: 24001695]
33. Centers for Disease Control and Prevention (CDC), 2018 Standardizing Hormone Measurements Program, Web site. Available at: [https://www.cdc.gov/labstandards/pdf/hs/HoSt\\_Brochure.pdf](https://www.cdc.gov/labstandards/pdf/hs/HoSt_Brochure.pdf). Accessed June 7, 2018.

**KEY POINTS**

- Rigorous approaches to harmonization and standardization of clinical assays have been published.
- Less-formal approaches to standardization can serve a useful purpose of improving harmonization of liquid chromatography–tandem mass spectrometry (LC-MS/MS) assays before the completion of formal harmonization projects.
- Factors that can affect the harmonization process are discussed with particular emphasis on LC-MS/MS protein assays.

**Fig. 1.**

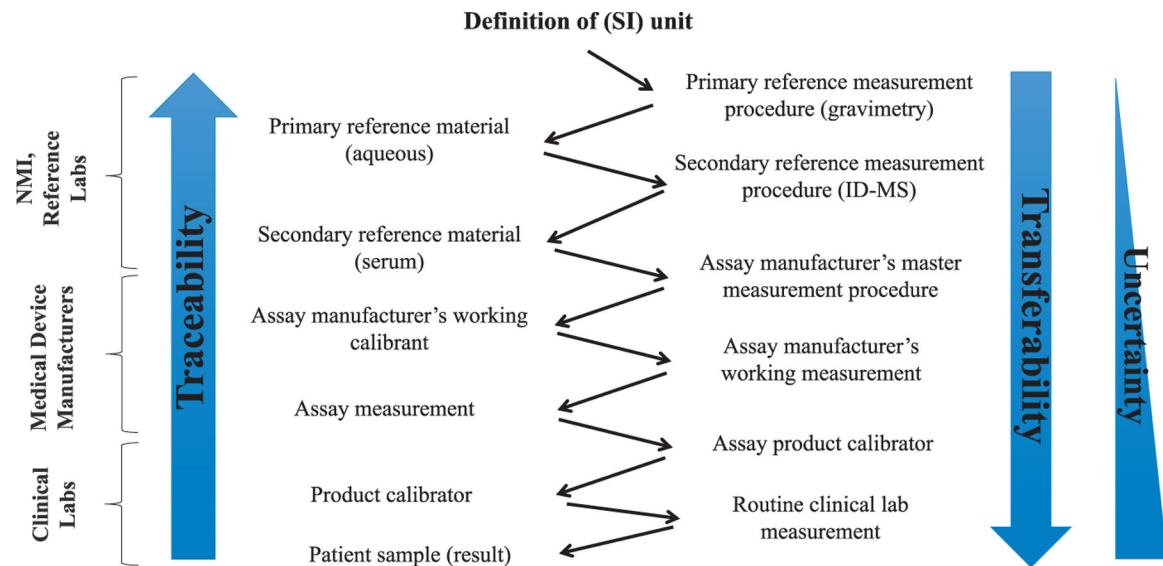
Overview of a general approach to manage harmonization of a measurand. IVD, in vitro diagnostic; JCTLM, Joint Committee for Traceability of Laboratory Medicine. Greg Miller W, Myers GL, Lou Gantzer M, et al. Roadmap for Harmonization of Clinical Laboratory Measurement Procedures Clinical Chemistry 2011;57(8):1108–17; *Reproduced with permission from* the American Association for Clinical Chemistry.



**Fig. 2.**

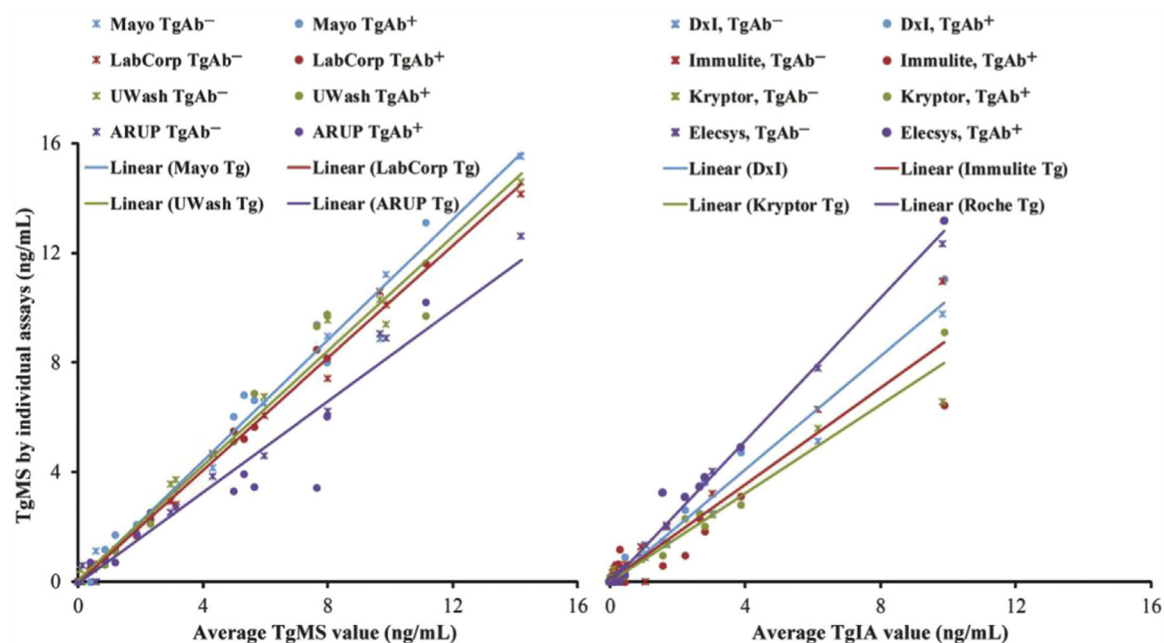
General process for assessing and achieving harmonization (equivalency) of clinical laboratory measurement results. Greg Miller W, Myers GL, Lou Gantzer M, et al. Roadmap for Harmonization of Clinical Laboratory Measurement Procedures Clinical Chemistry 2011;57(8):1108–17; *Reproduced with permission from the American Association for Clinical Chemistry.*





**Fig. 3.**

Illustration of traceability chain for standardizing methods to an agreed-on absolute standard of accuracy. NMI, National Measurement Institute. (*Data from* International Organization for Standardization. (2003–2008). In vitro diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials (ISO 17511:2003). Retrieved from: <https://www.iso.org/standard/30716.html>)



**Fig. 4.** Results from preliminary Tg harmonization study. (From Netzel BC, Grant RP, Hoofnagle AN. First steps toward harmonization of LC-MS/MS thyroglobulin assays: letter to the editor. Clin Chem 2015;62:1; with permission.)

		Digestion variability	PTMs	Protein folding	Pre-enrichment	Protein complex
Better calibrator → Better internal standard →	Full-length protein, mammalian expressed	✓	✓	✓	✓	✗
	Full-length protein, bacterial/yeast expressed	✓	✗	✓	✓	✗
	Concatamers	✓	✗	✗	✗	✗
	Synthetic [ $\pm$ SIL] peptides	✗	✗	✗	✗	✗
	chemical labeling (iTRAQ, TMT, O <sub>18</sub> )	N/A – relative quantification				
	cell culture, SILAC					

**Fig. 5.**

Example of a decision matrix for choosing calibrators. Green check marks indicate an analytical consideration for which the given calibrator accounts for a given source of bias. iTRAQ, Isobaric tags for relative and absolute quantitation; N/A, not applicable; 18O-Oxygen-18; PTM, Post translational modification; SILAC, Stable Isotope Labeling by/with Amino acids in Cell culture; SIL, stable isotope labeled; TMT, Tandem Mass Tag.